WHAT IS CLAIMED IS:

1. An automated method of designating text, taken from a set of citing documents, as reasons for citing (RFC) that are associated with respective citing instances of a cited document, the method comprising:

obtaining contexts of the citing instances in the respective citing documents, each context including text that includes the citing instance and text that is near the citing instance;

analyzing the content of the contexts; and

selecting, from the citing instances' context, text that constitutes the RFC, based on the analyzed content of the contexts.

2. An automated method of designating text, taken from a set of citing documents, as reasons for citing (RFC) associated with respective citing instances of a cited document, the method comprising:

inputting text from the citing documents;

dividing the citing documents' text to define paragraphs, and dividing the paragraphs to define sentences;

obtaining contexts of the citing instances in the respective citing documents, each context including: a sentence that includes the citing instance and at least one sentence that is near the citing instance;

generating a content word list based on words that are in the citing documents' contexts;

calculating, for the sentences in the citing documents' contexts, respective content scores that are based on frequency counts of content words that are recited in the respective sentences; and

selecting, from the citing documents' contexts, the sentences that constitute the RFC, based on the calculated content scores.

3. The method of claim 2, wherein the content word generating step includes:

generating the content word list based on words that are included in the contexts of at least two of the citing documents.

4. The method of claim 2, wherein the content word generating step includes:

generating the content word list based on words that are included both in the cited document itself and in the context of at least one citing document.

5. An automated method for selecting content words from documents, and for determining content scores for respective content words that indicate the content words' degree of relevance, the method comprising:

associating paragraphs from the documents;

processing text in the associated paragraphs to eliminate text that conveys little about the content of the paragraphs;

determining common words that are not eliminated by the processing step and that are found in plural paragraphs, while tallying content scores that indicate respective numbers of paragraphs in which the respective common words are encountered; and

forming the content word list as including the common words linked to respective content scores.

6. The method of claim 5, wherein the paragraph associating step includes:

pairing paragraphs from among documents that cite a cited document.

7. The method of claim 5, wherein the paragraph associating step includes:

pairing paragraphs from a cited document with paragraphs from documents that cite the cited document.

8. The method of claim 5, wherein the processing step includes:

removing from the paired paragraphs, noise words that convey little information about the content of the paragraphs; and

stemming the words of the paired paragraphs to a length that preserves their essential character while eliminating characters that convey little information about the word's identity.

9. An automated method of finding different morphological forms of a word, comprising:

inputting a word; and

stemming the word by eliminating any letters after the N^{th} letter from the beginning of the word, wherein N is a positive integer.

10. The method of claim 9, wherein N=6.

11. An automated method of scoring sentences in citing documents, to indicate relevance of content of the respective sentences to reasons that a cited document is cited, the method comprising:

calculating respective initial content scores (ICSs) for the sentences in the citing documents, based on the content of the sentences;

calculating respective distances (Ds) of the sentences in the citing documents from respective citing instances of the cited document; and

calculating respective content scores (CSs) for the sentences in the citing documents, based on at least the ICSs and the distances.

12. The method of claim 11, further comprising normalizing the ICSs to form normalized initial content scores (NICSs) for use by the CS calculation step, by taking into account:

a) numbers of words in the respective sentences; and

b) a largest frequency count in a content word list that includes:

1) a set of content words that are found in the sentences of the citing documents, and

2) a set of frequency counts linked to corresponding content words in the set of content words.

13. The method of claim 11, further comprising:

modifying the distances D to form respective modified absolute distances (MADs) for use by the CS calculation step, based on criteria relating to predetermined statistical observations of the implications of placement of a sentence in the citing document relative to the citing instance.

14. The method of claim 13, wherein the criteria include:

whether the sentence is in the same paragraph as the citing instance.

15. The method of claim 13, wherein the criteria include:

whether the sentence is located after the citing instance.

16. An apparatus for designating text, taken from a set of citing documents, as reasons for citing (RFC) that are associated with respective citing instances of a cited document, the apparatus comprising:

means for obtaining contexts of the citing instances in the respective citing documents, each context including text that includes the citing instance and text that is near the citing instance;

means for analyzing the content of the contexts; and

means for selecting, from the citing instances' context, text that constitutes the RFC, based on the analyzed content of the contexts.

17. An apparatus for designating text, taken from a set of citing documents, as reasons for citing (RFC) associated with respective citing instances of a cited document, the apparatus comprising:

means for dividing the citing documents' text to define paragraphs, and for dividing the paragraphs to define sentences;

means for obtaining contexts of the citing instances in the respective citing documents, each context including: a sentence that includes the citing instance and at least one sentence that is near the citing instance;

means for generating a content word list based on words that are in the citing documents' contexts;

means for calculating, for the sentences in the citing documents' contexts, respective content scores that are based on frequency counts of content words that are recited in the respective sentences; and

means for selecting, from the citing documents' contexts, the sentences that constitute the RFC, based on the calculated content scores.

18. The apparatus of claim 17, wherein the content word generating means includes:

means for generating the content word list based on words that are included in the contexts of at least two of the citing documents.

19. The apparatus of claim 17, wherein the content word generating means includes:

means for generating the content word list based on words that are included both in the cited document itself and in the context of at least one citing document.

20. An apparatus for selecting content words from documents, and for determining content scores for respective content words that indicate the content words' degree of relevance, the apparatus comprising:

means for associating paragraphs from the documents;

means for processing text in the associated paragraphs to eliminate text that conveys little about the content of the paragraphs;

means for determining common words that are not eliminated by the processing means and that are found in plural paragraphs, while tallying content scores that indicate respective numbers of paragraphs in which the respective common words are encountered; and

means for forming the content word list as including the common words linked to respective content scores.

21. The apparatus of claim 20, wherein the paragraph associating means includes:

means for pairing paragraphs from among documents that cite a cited document.

22. The apparatus of claim 20, wherein the paragraph associating means includes:

means for pairing paragraphs from a cited document with paragraphs from documents that cite the cited document.

23. The apparatus of claim 20, wherein the processing means includes:

means for removing from the paired paragraphs, noise words that convey little information about the content of the paragraphs; and

means for stemming the words of the paired paragraphs to a length that preserves their essential character while eliminating characters that convey little information about the word's identity.

24. An apparatus of finding different morphological forms of a word, the apparatus comprising:

means for stemming the word, the stemming means including:

means for eliminating any letters after the N^{th} letter from the beginning of the word;

wherein N is a positive integer.

25. The apparatus of claim 24, wherein N=6.

26. An apparatus of scoring sentences in citing documents, to indicate relevance of content of the respective sentences to reasons that a cited document is cited, the apparatus comprising:

means for calculating respective initial content scores (ICSs) for the sentences in the citing documents, based on the content of the sentences;

means for calculating respective distances (Ds) of the sentences in the citing documents from respective citing instances of the cited document; and

means for calculating respective content scores (CSs) for the sentences in the citing documents, based on at least the ICSs and the distances.

27. The apparatus of claim 26, further comprising means for normalizing the ICSs to form normalized initial content scores (NICSs) for use by the CS calculation means, by taking into account:

a) numbers of words in the respective sentences; and

b) a largest frequency count in a content word list that includes:

1) a set of content words that are found in the sentences of the citing documents, and

2) a set of frequency counts linked to corresponding content words in the set of content words.

28. The apparatus of claim 26, further comprising:

means for modifying the distances D to form respective modified absolute distances (MADs) for use by the CS calculation means, based on criteria relating to predetermined statistical observations of the implications of placement of a sentence in the citing document relative to the citing instance.

29. The apparatus of claim 28, wherein the criteria include:

whether the sentence is in the same paragraph as the citing instance.

30. The apparatus of claim 28, wherein the criteria include:

whether the sentence is located after the citing instance.

31. A computer-readable memory that, when used in conjunction with a computer, can carry out a method of designating text, taken from a set of citing documents, as reasons for citing (RFC) that are associated with respective citing instances of a cited document, the computer-readable memory comprising:

computer-readable code for obtaining contexts of the citing instances in the respective citing documents, each context including text that includes the citing instance and text that is near the citing instance;

computer-readable code for analyzing the content of the contexts; and

computer-readable code for selecting, from the citing instances' context, text that constitutes the RFC, based on the analyzed content of the contexts.

32. A computer-readable memory that, when used in conjunction with a computer, can carry out a method of designating text, taken from a set of citing documents, as reasons for citing (RFC) associated with respective citing instances of a cited document, the computer-readable memory comprising:

computer-readable code for inputting text from the citing documents;

computer-readable code for dividing the citing documents' text to define paragraphs, and dividing the paragraphs to define sentences;

computer-readable code for obtaining contexts of the citing instances in the respective citing documents, each context including: a sentence that includes the citing instance and at least one sentence that is near the citing instance;

computer-readable code for generating a content word list based on words that are in the citing documents' contexts;

computer-readable code for calculating, for the sentences in the citing documents' contexts, respective content scores that are based on frequency counts of content words that are recited in the respective sentences; and

computer-readable code for selecting, from the citing documents' contexts, the sentences that constitute the RFC, based on the calculated content scores.

33. The computer-readable memory of claim 32, wherein the content word generating computer-readable code includes:

computer-readable code for generating the content word list based on words that are included in the contexts of at least two of the citing documents.

34. The computer-readable memory of claim 32, wherein the content word generating computer-readable code includes:

computer-readable code for generating the content word list based on words that are included both in the cited document itself and in the context of at least one citing document.

35. A computer-readable memory that, when used in conjunction with a computer, can carry out an automated method for selecting content words from documents, and for determining content scores for respective content words that indicate the content words' degree of relevance, the computer-readable memory comprising:

computer-readable code for associating paragraphs from the documents;

computer-readable code for processing text in the associated paragraphs to eliminate text that conveys little about the content of the paragraphs;

computer-readable code for determining common words that are not eliminated by the processing computer-readable code and that are found in plural paragraphs, while tallying content scores that indicate respective numbers of paragraphs in which the respective common words are encountered; and

computer-readable code for forming the content word list as including the common words linked to respective content scores.

36. The computer-readable memory of claim 35, wherein the paragraph associating computer-readable code includes:

computer-readable code for pairing paragraphs from among documents that cite a cited document.

37. The computer-readable memory of claim 35, wherein the paragraph associating computer-readable code includes:

computer-readable code for pairing paragraphs from a cited document with paragraphs from documents that cite the cited document.

38. The computer-readable memory of claim 35, wherein the processing computer-readable code includes:

computer-readable code for removing from the paired paragraphs, noise words that convey little information about the content of the paragraphs; and

computer-readable code for stemming the words of the paired paragraphs to a length that preserves their essential character while eliminating characters that convey little information about the word's identity.

39. A computer-readable memory that, when used in conjunction with a computer, can carry out a method of finding different morphological forms of a word, the computer-readable memory comprising:

computer-readable code for stemming the word by eliminating any letters after the N^{th} letter from the beginning of the word, wherein N is a positive integer.

40. The computer-readable memory of claim 39, wherein N=6.

41. A computer-readable memory that, when used in conjunction with a computer, can carry out an automated method of scoring sentences in citing documents, to indicate relevance of content of the respective sentences to reasons that a cited document is cited, the computer-readable memory comprising:

computer-readable code for calculating respective initial content scores (ICSs) for the sentences in the citing documents, based on the content of the sentences;

computer-readable code for calculating respective distances (Ds) of the sentences in the citing documents from respective citing instances of the cited document; and

computer-readable code for calculating respective content scores (CSs) for the sentences in the citing documents, based on at least the ICSs and the distances.

42. The computer-readable memory of claim 41, further comprising computer-readable code for normalizing the ICSs to form normalized initial content scores (NICSs) for use by the CS calculation computer-readable code, by taking into account:

a) numbers of words in the respective sentences; and

b) a largest frequency count in a content word list that includes:

1) a set of content words that are found in the sentences of the citing documents, and

2) a set of frequency counts linked to corresponding content words in the set of content words.

43. The computer-readable memory of claim 41, further comprising:

computer-readable code for modifying the distances D to form respective modified absolute distances (MADs) for use by the CS calculation computer-readable code, based on criteria relating to predetermined statistical observations of the implications of placement of a sentence in the citing document relative to the citing instance.

44. The computer-readable memory of claim 43, wherein the criteria include:

whether the sentence is in the same paragraph as the citing instance.

45. The computer-readable memory of claim 43, wherein the criteria include:

whether the sentence is located after the citing instance.